THE COVARIANCE MATRIX OF NORMAL ORDER STATISTICS

BY

C. S. DAVIS and M. A. STEPHENS

TECHNICAL REPORT NO. 14
FEBRUARY 21, 1978

PREPARED UNDER GRANT
DAAG29-77-G-0031
FOR THE U.S. ARMY RESEARCH OFFICE

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government
Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE COVARIANCE MATRIX OF NORMAL ORDER STATISTICS

By

C.S. Davis and M.A. Stephens

TECHNICAL REPORT NO. 14

FEBRUARY 21, 1978

Prepared under Grant DAAG29-77-G-0031
For the U.S. Army Research Office
Herbert Solomon, Project Director

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Partially supported under Office of Naval Research Contract N00014-76-C-0475 (NR-042-267) and issued as Technical Report No. 254.
THE COVARIANCE MATRIX OF NORMAL ORDER STATISTICS

C.S. Davis and M.A. Stephens
McMaster University
Hamilton, Ontario, Canada

ABSTRACT
An approximation is given to calculate $V$, the covariance matrix for normal order statistics. The approximation gives considerable improvement over previous approximations, and the computing algorithm is available from the authors.

1. INTRODUCTION
Many statistical methods involve order statistics, and for a proper study of these methods the covariance matrix $V$ of a sample of order statistics is needed. For a few important distributions (e.g., the uniform and exponential), the entries $V_{ij}$ can be expressed in closed form and can be calculated easily; but for most parent populations each $V_{ij}$ involves a double integral, so that accurate tabulation is difficult and expensive. In particular, for the normal population, $V$ has so far been published only for samples of size $n \leq 20$, (see e.g., Sarhan and Greenberg, 1956; Owen, 1962). The need for good tables of $V$, for
many populations, was pointed out by Hastings et al. (1947) and
the magnitude of the problem of exact calculation was also
stressed; subsequently, series expansions for $V_{ij}$ have been given
by Plackett (1958) and by David and Johnson (1954). Saw (1960)
compared these expansions and concluded that although Plackett's
series converges a little faster for a normal population, there
were computational advantages in the David-Johnson method. The
David-Johnson formulae give $V_{ij}$ up to terms in $(n+2)^{-3}$.

In this paper we are concerned with $V$ for normal order
statistics. We give a technique by which one can obtain an
excellent approximation for $V$, by starting with the values given
by the David-Johnson formulae and modifying them by use of cer-
tain identities and specially tabulated values for normal order
statistics.

Suppose $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the order statistics (in
ascending order) of a sample of size $n$ from a normal distribution
with mean 0 and variance 1; let $m_i = E(X_{(i)})$, where $E$ stands for
expectation, and let $V$ have entries $V_{ij} = E(X_{(i)} - m_i)(X_{(j)} - m_j)$.

Three useful identities are:

1. For any $i$, $\sum_{j=1}^{n} V_{ij} = 1$ ;

$$E(X_{(1)}^2) = E(X_{(1)}X_{(2)}) + 1 ;$$

2. The trace of $V$ is $\text{tr}(V) = n - \sum_{i=1}^{n} m_i^2$ .

From (2) we obtain

$$V_{12} = E(X_{(1)})^2 - m_1 m_2 - 1 .$$

Of these, (1) is very well known, (2) is given in, e.g.,
Govindarajulu (1963), and (3) is easily proved as follows:
\[ \text{tr}(V) = \sum_{i} \left( X_{(i)} - m_{i}\right)^2 = \sum_{i} \left( X_{(i)} \right)^2 - \sum_{i} m_{i}^2 = n - \sum_{i} m_{i}^2. \]

We shall also need the results, obtained from the symmetry of \( V \):

\[ V_{ij} = V_{ji} = V_{rs} = V_{sr}; \quad r = n+1-i, \quad s = n+1-j. \] (5)

Values of \( m_{i} \) have been extensively tabulated; e.g., for \( n \leq 20 \), to 10 decimal places (d.p.) in Teichroew (1956), and, for all \( n \leq 100 \) and at intervals for \( n \leq 400 \), to 5 d.p. in Harter (1961). The sum \( \sum m_{i}^2 \) is given for \( n \leq 100 \), to 5 d.p. in Pearson and Hartley (1972, Table 13) and in Owen (1962, p. 154). Ruben (1954) examines the distributions of \( m_{i} \) from a geometric viewpoint; among the results in his paper he gives moments of the extreme order statistic and tabulates the variance of \( X_{(1)} \) i.e., \( V_{11} \), for \( n \leq 50 \), to 8 d.p. Borenius (1966) has extended this tabulation to \( n \leq 120 \). These exact values are important in obtaining a good approximation for \( V \), since \( V_{11} \) is the most inaccurate term in David and Johnson's formulae. LaBrecque (1973) used the David and Johnson technique, and the correct \( V_{11} \) to calculate certain functions of \( m' = (m_{1}, m_{2}, ..., m_{n}) \) and \( V \). In the next section we use \( V_{11} \) and the other identities above to give a considerable improvement over the David-Johnson formulae used alone.

By normalization of a row we shall mean keeping certain terms fixed and then multiplying the others by a constant, to ensure that the sum of all elements is 1, as required by (1) above.

2. AN APPROXIMATION FOR V

The calculations for \( V \) follow the following steps:

(a) Insert the correct \( V_{11} \) from the tables referenced above;
(b) Insert the correct \( V_{12} \) from (4);
(c) Insert the rest of row 1 using the David and Johnson formulae;
(d) Keep $V_{11}$ and $V_{12}$ fixed, and normalize row 1.  When row 1 has been calculated, fill in column 1, row n and column n from the symmetry relations (5).

(e) Apart from terms already calculated from steps (a) through (d) (i.e., $V_{21}$ and $V_{2n}$), calculate row 2 from the David and Johnson formulae, and normalize row 2.  Fill in column 2 and row $n-1$ and column $n-1$ from the symmetry of $V$.  Continue with successive rows until all rows are normalized.

These operations make the top left corner of $V$ correct, and the rows more accurate than before; but the trace will not satisfy (3).  This identity can be used to give further improvement as follows:

(f) Change $V_{22}$ and its equal, $V_{pp}$ ($p = n-1$), so that (3) is satisfied; then renormalize row 2 with $V_{21}$, $V_{22}$, $V_{2n}$ fixed.  Fill in symmetric terms, in columns 2 and $n-1$ and row $n-1$.

(g) Renormalize successively all rows as for row 2; i.e., leave fixed the diagonal term and terms calculated from symmetry relations with previous rows.  The entire matrix $V$ will be complete when row $n/2$ is renormalized, for $n$ even, or row $(n-1)/2$, for $n$ odd.  In the latter case, $(n$ odd), the middle row will not satisfy (1).  The procedure could be iterated to improve this, but our experience suggests that this is not necessary.

3. ACCURACY OF THE METHOD

When the David-Johnson formulae are used alone, by far the greatest error, for those values of $n$ ($< 20$) for which comparisons can be made over the entire matrix $V$, occurs at $V_{11}$.  For this particular entry we can, of course, extend comparisons to $n = 120$; the error is about 0.00440 at $n = 20$ (about 1.6%) and diminishes very slowly to 0.00395 at $n = 120$ (about 2.2%).  In our computations we used the algorithm of Cunningham (1969) to give the inverse of the normal distribution; this will give computational errors much smaller than those in the approximation itself.  The very slow decrease lends support to misgivings
expressed by David and Johnson on the convergence properties, for extreme values, of their series. Comparisons of other terms are in the Table; we have selected those terms where either the

<table>
<thead>
<tr>
<th>Approximation:</th>
<th>D-J</th>
<th>D-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Element</td>
<td>True Value</td>
</tr>
<tr>
<td>10</td>
<td>$V_{23}$</td>
<td>.146623</td>
</tr>
<tr>
<td>10</td>
<td>$V_{33}$</td>
<td>.175003</td>
</tr>
<tr>
<td>15</td>
<td>$V_{22}$</td>
<td>.179122</td>
</tr>
<tr>
<td>18</td>
<td>$V_{13}$</td>
<td>.094617</td>
</tr>
<tr>
<td>18</td>
<td>$V_{22}$</td>
<td>.166293</td>
</tr>
<tr>
<td>20</td>
<td>$V_{22}$</td>
<td>.159573</td>
</tr>
</tbody>
</table>

David-Johnson formulae used alone give largest error (omitting $V_{11}$) or where the new technique gives largest error; for $n = 15$ and $n = 20$ these both occur at $V_{22}$. The new method reduces the maximum error to about one-fifth its previous value. From a percentage point of view, the new approximation gives largest percentage error at $V_{1n}$, where the true covariance is smallest; this maximum percentage error is of the order of 0.15%; the maximum absolute error is generally less than 0.05%. A comparison of the relative sizes of the errors in the Table, and those quoted above for $V_{11}$ shows how important it is to have the exact values for $V_{11}$ to make a good start in approximating $V$. We have suggested the upper limit $n = 120$ because exact values are known to this point. However, $V_{11}$ approaches zero like $1/(\ln n)$; in fact, asymptotically $V_{11} \ln n$ has limit $\pi^2/12 = 0.82$ (Cramér, 1946, p. 376), though $V_{11} = 0.85/(\ln n)$ gives a more accurate approximation in the region $110 \leq n \leq 120$ (error less than 0.0001). Thus, the use of the algorithm could be extended. In order to
use the David-Johnson formulae, a computer would be needed; the steps given above can be very easily programmed and it seems worthwhile to get the extra accuracy, particularly if, as in some applications, the inverse of $V$ is required. A Fortran program for the entire procedure is available from the authors.

References


THE COVARIANCE MATRIX OF NORMAL ORDER STATISTICS

C.S. Davis and M.A. Stephens

Department of Statistics
Stanford University
Stanford, CA 94305

U.S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

February 21, 1978

Approved for Public Release; Distribution Unlimited.

Normal distribution, covariance matrix calculations, normal order statistics, generalized least squares.

An approximation is given to calculate \( V \), the covariance matrix for normal order statistics. The approximation gives considerable improvement over previous approximations, and the computing algorithm is available from the authors.